

Deploying the eCAS to Operation

John J. Tran

April 14, 2009

Agenda

Overview of the Curation Process

CurationSteps

Pre-curation detective work

Background Information

Leveraging existing metadata

Curation

Files to edit

Ingestion

Ingestion Scripts

Manual Curation Steps

1. Pre-curation detective work
2. Update policy, metadata, and configuration files
3. Do the ingestion

Explicit Info

Before you can begin to curate data, some explicit information on the product are required:

- ▶ Dataset description
- ▶ Principle PI and data custodian
- ▶ Institution Information – short and long name
- ▶ Organ
- ▶ Instrument

Implicit Info

Some additional implicit information required:

- ▶ Number of products to be ingested
- ▶ Nature of data being ingested

It is often the case that existing metadata might have already been created to curate a similar product, in which case you should leverage this by simply copying the contents of the policy folder.

Policy Files

The curation process involves editing three files in the policy folder:

1. `elements.xml` – update the change the description comment field
2. `product-type-element-map.xml` – note the full hierarchical name URN
3. `product-types.xml`

The easiest way to get started is to copy an existing dataset from the policy folder then modify the above files.

What is a URN?

A URN is a unique identifier specific to the data product being curated.

Extraction Files

You will need to modify (create) two extraction files:

- ▶ `.met` – metadata extraction file specific to this ingestion instance
- ▶ `.conf` – configuration file (for the crawler), usually references the above metadata extraction file

Using the existing ingestion script

We currently have an existing extraction:

```
copy_and_rewrite_ingest
```

This script will call the copy and rewrite extractor as part of the ingestion process